Second Order Properties of Generalized Impulse Response Functions *

Simon M. Potter †

July 1998
(Preliminary/Incomplete)

*I would like to thank three anonymous referees, Patrick Asea, Joe Ostroy and Bill Zame for helpful suggestions and comments. Research supported in part by the National Science and the Academic Senate UCLA.

†Research and Markets Analysis Group, Federal Reserve Bank of New York, 33 Liberty St, New York, NY 10045-0001. email: simon.potter@ny.frb.org

1 INTRODUCTION

In the last few years there has been a growing interest in the use of nonlinear models to describe economic time series. Some of the standard tools of analysis of linear time series models have been extended to nonlinear models such as the impulse response function (see Gallant, Rossi and Tauchen, 1993). This allows researchers to examine asymmetric responses to shocks depending on the state of the system and the size and sign of the shock itself. However, there is some ambiguity as to how to provide measures of persistence* that average across the effect of shocks and the state of the system. One of the main contributions of this paper is to show that such average properties can be found directly from the autocovariance function of the time series itself without having to consider the nonlinear structure directly.

In the statistics literature, Priestley (1988) has developed a generalized transfer function analysis of nonlinear time series in the frequency domain. He notes that there is a 'non-uniqueness' property to the generalized transfer function of nonlinear time series. This stands in contrast to the uniqueness of generalized transfer functions of linear time series and the uniqueness of transfer functions defined by the autocovariance function of a time series (linear or nonlinear). This 'non-uniqueness' is caused by the shock and state dependence of the effect of shocks on a nonlinear time series.

This paper uses a generalized impulse response function (GI hereafter) in the time domain to further illuminate this 'non-uniqueness' property. The GI is originally developed by Potter (1995) for the univariate time series case and Koop, Pesaran and Potter (1996) in the multiple time series case.† In Potter (1995) the GI is defined to be a random variable given by the difference between the conditional expectation of the time series at time t and the conditional expection of the time series at time $t-1$. Since the GI is defined directly on the underlying probability space of the time series, nonlinear time series do have unique 'impulse response functions' in a suitably defined space of random variables.

In the previous work only the weak restriction of integrability was placed on the time series to define the GI. By the use of stochastic dominance concepts it was shown how to provide measures of persistence. However, these measures were not complete because of the partial ordering produced by stochastic dominance. In this paper I examine time series with the tighter restriction that their GI lie in a space where all infinite sequences of square integrable random variables are square summable. This restriction defines a Hilbert Space for the sequence of GI. In contrast, the linear impulse response functions/transfer functions (derived from the Wold Representation) of both linear and nonlinear covariance stationary time series are in the Hilbert space of square summable sequences. It is shown that covariance stationarity combined with an assumption of strict stationarity is sufficient for time series to have GI that are square integrable and square summable.

One can 'map' the GI into the space of autocovariance functions using an expectations operator conditioned on past values of the time series. This establishes that the dynamic responses contained in one nonlinear time series model can be interpreted as the dynamic responses of many different linear time series models. Further, it is established that the average of the autocovariance functions produced in this manner is the autocovariance function of the time series. Hence, certain average features of the dynamics of nonlinear time series models can be obtained directly from their autocovariance functions.In particular attention is placed on the family of spectral density functions obtained from the autocovariance functions generated by the GI. This allows one to measure how nonlinearity varies across the frequency domain and gives a measure of the total nonlinearity in the time series. .

The outline of the rest of the paper is as follows. Section 2 gives the definition of the GI. Section 3 constructs a Hilbert space for the GI and develops a mapping into the space of autocovariance functions. Section 4 constructs a measure of nonlinearity in the frequency domain and gives an overall measure of nonlinearity in a time series. Section 5 contains an example of the use of the second order properties of GI to analyze the behavior of a nonlinear moving average. Section 6 discusses some computational techniques. Section 7 applies

*Persistence is to be interpreted as the size of the response at a particular horizon.

†Gallant, Rossi and Tauchen (1993) use a different definition that is not amenable to the analysis presented in this paper.

the techniques to Hamilton's (1989) and Potter's (1995) models of US GNP. An appendix contains proofs of results in the main text.

NOTATION AND CONVENTIONS

Throughout the paper the convention will be maintained that lower case letters refer to a realization of a random variable and upper case letters to the random variable. The time series, $\{Y_t\}$, examined are all assumed to be defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. In order to clarify the notation with respect to the expectations operator the double bar convention of Billingsley (1986) will be used. Thus, $E[Y_{t+n}||\mathcal{G}]$ will mean the random variable called the conditional expected value of Y_{t+n} given that $\mathcal{G}$ is a $\sigma-$field in $\mathcal{F}$. $E[Y_{t+n}|G]$ will signify the predicted value of Y_{t+n} given the realized information in G.

Below we will be concerned with time series and the sequence of increasing Borel $\sigma-$fields,

$$\mathcal{F}_t = \sigma(\ldots, Y_{-1}, Y_0, Y_1, \ldots, Y_t).$$

The convention will be maintained that $y^t(\omega)$ identifies an individual realization of $(y_t, \ldots, y_1, y_0, y_{-1}, \ldots)$. I restrict attention to the case of strictly stationary time series with the exception that nonstationarity due to deterministic components will be directly allowed for.

2 GENERALIZED IMPULSE RESPONSE FUNCTIONS

Intuitively the GI is the difference between two conditional expectations where the conditioning information differs by one period. In order to formally define the GI the following adaption of a theorem from Billingsley(1986) is useful.

Theorem 33.3: Billingsley (1986)

If Y_t is a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{F}_t$ is a sequence of $\sigma-$fields in $\mathcal{F}$. Then there exists a sequence of functions $\mu_n(t, H, \omega)$ defined for H in the Borel Sets of the line $\mathcal{B}(R)$, and ω in Ω with these two properties:

1. *For each ω in Ω $\mu_n(t, H, \omega)$ is, as a function of H, a probability measure on R.*
2. *For each H in $\mathcal{B}(R)$, $\mu_n(t, H, \omega)$ is, as a function of ω, a version of $P[Y_{t+n} \in H||\mathcal{F}_t]_\omega$.*

Using this result one can define the GI by:

$$GI_n(t, \omega) = \int y\mu_n(t, dy, \omega) - \int y\mu_{n+1}(t-1, dy, \omega).$$

The properties of the random variable GI_n can be used to provide measures of the 'distance' between the conditional distributions, $\mu_n(t, H, \omega)$ and $\mu_{n+1}(t-1, H, \omega)$ for a wide class of time series. It is also possible to analyze various information theoretic measures of the distance between the two distribution functions as described in Tong (1995).

The integral of GI_n against the underlying probability measure:

$$\int GI_n(t, \omega) dP,$$

will return the value of zero for all n.. Further, under the condition that the underlying non-deterministic component of the time series is ergodic, the limiting behavior of GI_n will be zero with probability one. In general, a time series $\{Y_t\}$ will have a well-defined $\{GI_n\}$ if $\{W_t\}$ is in $L^1(\Omega, \mathcal{F}, \mathbf{P})$, where

$$W_t = Y_t - E[Y_t||\mathcal{F}_{-\infty}].$$

That is, the purely deterministic (in the nonlinear sense) component of Y_t is removed by the operation of constructing the GI.[‡] Here the stronger requirement of square integrability on the non-deterministic component will be imposed to develop the relationship to the usual second order analysis.

I restrict attention to time series where $\mathcal{F}_t = \sigma(V_t) \bigcup \mathcal{F}_{t-1}$ and $\{V_t\}$ is a martingale difference sequence with respect to $\mathcal{F}_t$. Allowing for a slight abuse of notation, define:

$$E[Y_{t+n}||V_t, \mathcal{F}_{t-1}] = E[Y_{t+n}||\mathcal{F}_t],$$

and adopt the convention that V_t is the 'shock' and $\mathcal{F}_{t-1}$ is the 'history'. We will also use the same notation for the GI

$$GI_n(V_t, \mathcal{F}_{t-1}) = E[Y_{t+n}||V_t, \mathcal{F}_{t-1}] - E[Y_{t+n}||\mathcal{F}_{t-1}].$$

I signify a realization of the GI by:

$$GI_n(v_t, y^{t-1}(\omega)) = E[Y_{t+n}|v_t, y^{t-1}(\omega)] - E[Y_{t+n}|y^{t-1}(\omega)],$$

3 A Hilbert Space for Generalized Impulse Response Functions

In order to gain some intuition for the approach that will be followed consider the GI of a purely non-deterministic linear model with IID innovations:

$$GI_n(V_t, \mathcal{F}_{t-1}) = \psi_n V_t.$$

Note that if the linear model is covariance stationary then $E[V_t^2] = \sigma^2 < \infty, \sum \psi_n^2 < \infty$ or equivalently:

$$\sum_{n \geq 0} E[GI_n(V_t, \mathcal{F}_{t-1})^2] < \infty.$$

The autocovariance function of the time series is given by:

$$\gamma(k) = \sigma_v^2 \sum_{n \geq 0} \psi_n \psi_{n+k} = \sum_{n=-\infty}^{\infty} E\left[GI_n(V_t, \mathcal{F}_{t-1}) GI_{n+k}(V_t, \mathcal{F}_{t-1})\right].$$

The approach will be to mimic this operation for a generic generalized impulse response function. Start by considering the complete inner product space $\mathcal{A}(\Omega, \mathcal{F}, \mathbf{P}) = \ell^2\{L^2(\Omega, \mathcal{F}, \mathbf{P})\}$ of square summable (doubly) infinite sequences of square integrable random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ as the time series.

$$\mathcal{A} = \{Z_n, n = 0, \pm 1, \pm 2 : E[\sum_{n=-\infty}^{\infty} Z_n^2] < \infty\},$$

and inner product,

$$\langle \mathbf{Z}, \mathbf{X} \rangle = E[\sum_{n=-\infty}^{\infty} Z_n X_n].$$

[‡]This is non-deterministic in the nonlinear sense which is different from the linear one used in the Wold decomposition. For example, certain chaotic processes can be shown to have the same spectrum as white noise and no deterministic components from the linear perspective. However, allowing for nonlinear prediction would imply that the series only had a deterministic component.

The operations involved in the construction of this Hilbert Space are known as the 'direct sum' of the individual Hilbert Spaces, Dunford and Schwartz (1957 p.256-257). In our case the individual Hilbert spaces are all $L^2(\Omega, \mathcal{F}, \mathbf{P})$. Note that elements of $\mathcal{A}$ are stochastic processes which I index by n to avoid confusion with the time series itself which is indexed by t. Obviously stationary time series are not themselves elements of $\mathcal{A}$ since as $n \to \infty$ $E\left[Z_n^2\right]$ goes to zero.

Consider purely non-deterministic covariance stationary time series. By the Wold Representation we know that they all can be written in the following form:

$$Y_t = \sum_{n=0}^{\infty} \varphi_n U_{t-n},$$

where $\{U_t\}$ is an uncorrelated sequence of random variables, φ_0 is normalized to be 1 and $\sum_{n=0}^{\infty} \varphi_n^2 < \infty$. Hence, the sequence $\{\varphi_n U_{t-n}\}$ is in $\mathcal{A}$, if we assume for $n < 0$ $\varphi_n = 0$. Theorem 1 shows that with the added restriction of strict stationarity the GI of covariance stationary time series are in $\mathcal{A}$. The requirement of strict stationarity is required to ensure that the form of the conditional expectation function does not vary through time.

Theorem 1 *All covariance (strictly) stationary time series have GI that are in $\mathcal{A}$.*

Once again since the GI purges the time series of any deterministic components a wide range of time series will have GI in $\mathcal{A}$.

Define an operator:

$$CE(\mathcal{F}_{t-1}) : \mathcal{A} \to \mathcal{A}_{\mathcal{F}_{t-1}},$$

by

$$\{\sum_{n=-\infty}^{\infty} E_{y^{t-1}(\omega)}\left[GI_{n+j} GI_{n+k}\right] : j, k = 0, \pm 1, \pm 2, \ldots\},$$

where

$$E_{y^{t-1}(\omega)}[GI_{n+j} GI_{n+k}] = \int_{y^{t-1}(\omega)} GI_{n+j} GI_{n+k} dP, \;\; E_{y^{t-1}(\omega)}[GI_{n+j}] = 0.$$

Note that this operator was 'applied' at the start of this section to a linear time series and the resulting object was the autocovariance of the time series itself. Theorem 1 shows that this result extends to the nonlinear time series case in the sense that the operator produces a stochastic process whose realizations are autocovariance functions.

Theorem 2

$\mathcal{A}_{\mathcal{F}_{t-1}}$ is the space of stochastic processes defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with realizations that are autocovariance functions of covariance stationary time series.

Define:

$$c_{y^{t-1}(\omega)}(j-k) = \sum_{n=-\infty}^{\infty} E_{y^{t-1}(\omega)}\left[GI_{n+j} GI_{n+k}\right].$$

The random variable corresponding to $c_{y^{t-1}(\omega)}(k)$ will be denoted by $C_{\mathcal{F}_{t-1}}(k)$. Similarly, to the GI this is a random variable on the probability space of the time series. Note that $CE(\mathcal{F}_{t-1})$ maps the GI of linear (homoscedastic) time series into a space $\mathcal{A}_{\mathcal{F}_{t-1}}$ with the same autocovariance function for all $y^{t-1}(\omega)$ that is equal to the autocovariance function of the time series itself (see the example at the start of the section). In contrast the GI of nonlinear time series are mapped to different autocovariance functions in $\mathcal{A}_{\mathcal{F}_{t-1}}$ for each

$y^{t-1}(\omega)$. However, it is possible to directly characterize the average of the autocovariance functions produced by the GI of the nonlinear time series.

Corollary to Theorem 2

If Y_t is a covariance stationary time series, with autocovariance function $\{\gamma(k)\}$ then

$$E[C_{\mathcal{F}_{t-1}}(k)] = \gamma(k), k = 0, 1, \ldots$$

The corollary suggests that among the wide range of possible definitions for impulse response functions of nonlinear time series the GI when combined with the operator above has the attractive property of being related to the autocovariance function of the time series. Further, amongst possible averaging operations applied to the GI this is the only one that when averaged returns the autocovarance function of the underlying time series and maps the GI of linear (homoscedastic) time series into one autocovariance function. One could apply a conditional expectation to the inner product that averaged over histories rather than shocks but this would mean that the GI of linear (homoscedastic) time series mapped to many autocovariance functions depending on the realization of V_t .

The relationship between the GI of a nonlinear time series (conditional on a history and integrating out over a shock) and autocovariance functions is particularly useful since linear operations applied to the autocovariance function will preserve the relationship in the corollary to Theorem 1.

It is possible to use Herglotz's Theorem (see Brockwell and Davis (1987)) to define a spectral distribution function from the autocovariance function $\{c_{y^{t-1}(\omega)}(k)\}$ under the condition that a time series has its GI in $\mathcal{A}$. However, it is more useful to consider the case where the autocovariance function of the GI of the time series, produced by the operations defined above, is absolutely summable, $\sum_k |C_{\mathcal{F}_{t-1}}(k)| < \infty$ with probability one. Obviously similar arguments to those used above to prove that covariance stationary time series have GI in $\mathcal{A}$ also show that time series with absolutely summable autocovariance functions will have GI with this property.

Under absolute summability, the discrete Fourier transform of $c_{y^{t-1}(\omega)}(k)$, $f_{y^{t-1}(\omega)}(\lambda)$, is defined by:

$$2\pi f_{y^{t-1}(\omega)}(\lambda) = c_{y^{t-1}(\omega)}(0) + 2\sum_{k=1}^{\infty} c_{y^{t-1}(\omega)}(k)\cos(\lambda k), \text{ for } 0 \leq \lambda \leq \pi.$$

Clearly, by the above corollary the average value of this discrete Fourier transform can be evaluated by considering the autocovariance function of the time series. Thus, an average measure of long run persistence for a nonlinear time series is given by the Fourier transformed evaluated at zero, i.e, the sum of the autocovariance function of the first difference of the series. On the other hand, the impulse response function derived from the Wold Representation of the time series will not be the average of impulse response functions defined for each fixed $y^{t-1}(\omega)$ unless the time series is linear since the moving average weights in the Wold Representation are not obtained by linear operations on the autocovariance functions. This can be seen most clearly by considering an expression for the moving average coefficients $\{\phi_n\}$ in terms of the Fourier transform of the series (see Brillinger, 1981 p79):

$$\phi_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(\frac{1}{2}\eta(0) + \sum_{l}^{\infty} \eta(l)\exp(-il\lambda))\exp(in\lambda)d\lambda,$$

where

$$\eta(l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(f(\lambda))\exp(il\lambda)d\lambda.$$

In the literature on persistence for integrated linear time series models there was some discussion of the relative merits of measuring persistence using the spectral density at zero vs the sum of the coefficients in the Wold Representation following Cochrane (1988). For nonlinear time series the sum of the autocorrelation function will not be equal to the average of the autocorrelation functions produced by the GI since

$$\frac{\gamma(k)}{\gamma(0)} \neq E[\frac{C_{\mathcal{F}_{t-1}}(k)}{C_{\mathcal{F}_{t-1}}(0)}].$$

Thus, there is the additional issue that the measure of persistence defined from the sum of the autocovariance function is not normalized by the size of the innovation variance. In terms of comparing models for the same time series this is not a major problem. An alternative approach would be to construct a 'Wold Representation' for the autocovariance functions produced by the GI using the innovations algorithm (see Brockwell and Davies, 1987 pp 165-6). The collection of Wold Representations could then be averaged to produce an overall measure of persistence that was more comparable across time series.

4 Measures of Nonlinearity

In the previous section it was shown that a defining characteristic of a linear homoscedastic time series was that applying the operator $CE(\mathcal{F}_{t-1})$ to its GI produces only one autocovariance function, that of the time series itself. In this section I consider various ways of measuring by how much a nonlinear time series model diverges from this feature of a linear time series.

One might think that nonlinearity could be more directly measured by considering the ratio

$$|\frac{GI_n(V_t,\mathcal{F}_{t-1})}{GI_o(V_t,\mathcal{F}_{t-1})}|,$$

which is a constant for each n for linear models. However, as discussed in Potter (1995) for nonlinear models it is possible for $GI_o(V_t,\mathcal{F}_{t-1})$ to be zero but $GI_n(V_t,\mathcal{F}_{t-1})$ to be non-zero implying that the ratio could not be used in all cases.

Consider a linear time series model with heteroscedastic innovations generated by for example an ARCH or GARCH process. The GI of this time series will be mapped by $CE(\mathcal{F}_{t-1})$ to many autocovariance functions. All of the autocovariance functions will have the same 'shape' but the scaling will be different. Alternatively, the autocovariance functions will all map to the same autocorrelation function when each is normalized by $c_{y^{t-1}(\omega)}(0)$. This suggests that if one is interested in measuring nonlinearity in the mean then the autocorrelation function is the appropriate measure. For example, one could measure the variance of the autocorrelation functions at a particular lag. Further, since the autocorrelation function is scale invariant it has the additional property that the amount of nonlinearity would not depend on the caling of the time series. However, since the autocorrelation function is defined on the set of integers it is not immediately obvious how to weight the contribution of different lags. Instead I consider using the spectral density function.

Using standard results one can define a (conditional) spectral density function, $s(\lambda|y^{t-1}(\omega))$ by $2 \times f_{y^{t-1}(\omega)}(\lambda)/c_{y^{t-1}(\omega)}(0)$ with the property that it integrates to 1 over $[0, \pi]$.[§] From the spectral representation of a time series we know that the height of the spectral density of the time series at frequency λ is a measure of the relative contribution of the random coefficient corresponding to this frequency in the spectral representation. A natural concept of the amount of nonlinearity in a time series is how much this contribution varies. One can measure the divergence from nonlinearity in the frequency domain using

$$NL(\lambda) = -\int_{y^{t-1}(\varpi)\in\mathcal{F}_{t-1}} \ln(\frac{s(\lambda|y^{t-1}(\omega))}{s(\lambda)})\, dP.$$

[§] As pointed out by a referee λ itself is not a random variable.

Notice that for a linear time series even with conditionally heteroskedastic innovations $s(\lambda|y^{t-1}(\omega)) = s(\lambda)$ and the measure is equal to zero. On the other hand, from above we know that the spectral densities from the GI do not average to the spectral density of the time series, hence it is possible that

$$-\int_{y^{t-1}(\varpi)\in\mathcal{F}_{t-1}} \ln(\frac{f(\lambda|y^{t-1}(\omega))}{f(\lambda)})\ dP \leq -\int_{y^{t-1}(\varpi)\in\mathcal{F}_{t-1}} \ln(\frac{c_{y^{t-1}(\omega)}(0)}{\gamma(0)})\ dP,$$

and thus for some frequencies the measure would be less than zero. Despite the fact that the measure might by negative for some frequencies an overall measure of nonlinearity in the time series defined by the integral:

$$TNL = \int_0^\pi NL(\lambda)s(\lambda)\ d\lambda,$$

would be positive. Since by simple manipulations one can show that TNL is the Kullback-Leibler distance between the joint distribution of $(y^{t-1}(\omega), \lambda)$ and the product of their marginals measured by the product of the marginals. To avoid the difficulties of NL(λ) being less than one I consider replacing s(λ) with

$$\overline{s(\lambda)} = \int_{y^{t-1}(\varpi)\in\mathcal{F}_{t-1}} s(\lambda|y^{t-1}(\omega))dP,$$

the average value of the spectral densities obtained from the GI in the definitions of NL and TNL.

The interpretation of NL(λ) is that high values represent frequencies for which knowledge of $y^{t-1}(\omega)$ provides information about the spectral density from the GI different from that contained in the average of the spectral densities from the GI.

5 Nonlinear Moving Average Examples

. In order to illustrate the use of the concepts without the need for complex computation I consider a nonlinear moving average in the spirit of Priestley (1988). Consider the nonlinear moving average:

$$Y_t = V_t + \theta V_{t-1}V_{t-2}^2,$$

where $V_t \sim IIDN(0,1)$. As in Priestley (1988) $\mathcal{F}_t$ should be defined to include the past values of the shocks. This time series is first order moving average from the covariance perspective. Its GI are given by:

$$GI_0 = V_t,$$

$$GI_1 = \theta V_t V_{t-1}^2,$$

$$GI_n = 0\ \forall n > 1.$$

The autocovariance function of the GI is given by:

$$C_{\mathcal{F}_{t-1}}(0) = 1 + \theta^2 V_{t-1}^4,$$

$$C_{\mathcal{F}_{t-1}}(1) = \theta V_{t-1}^2.$$

$$C_{\mathcal{F}_{t-1}}(k) = 0,\ \forall k > 1.$$

Standard calculations indicate that $\{C_{\mathcal{F}_{t-1}}(k)\}$ averages to the autocovariance function of the time series. The average of the discrete Fourier transform of $\{C_{\mathcal{F}_{t-1}}(k)\}$ at zero is given by:

$$2\pi E[f_{\mathcal{F}_{t-1}}(0)] = 1 + 3\theta^2 + 2\theta.$$

This expression is minimized at $\theta = -1/3$ with a value of $2/3$ but note that there is substantial variation around this value. Evaluating the discrete Fourier transform of $\{C_{\mathcal{F}_{t-1}}(k)\}$ at zero with $\theta = -1/3$ gives:

$$2\pi f_{\mathcal{F}_{t-1}}(0) = 1 + 1/9V_{t-1}^4 - 2/3V_{t-1}^2.$$

In contrast, if we restrict attention to the unit interval the average of the Fourier transform at zero is maximized at $\theta = 1$ with a value of 6. Evaluating the discrete Fourier transform of $\{C_{\mathcal{F}_{t-1}}(k)\}$ at zero with $\theta = 1$ gives:

$$2\pi f_{\mathcal{F}_{t-1}}(0) = 1 + V_{t-1}^4 + 2V_{t-1}^2,$$

Figure 1 shows the density of the discrete Fourier transform at frequency zero for the two cases, $\theta = -1/3, 1$. Note that dispersion around the average value of $2/3$ when $\theta = -1/3$ is substantial and asymmetric. The discrete Fourier transform for $\theta = 1$ first order stochastically dominates that for $\theta = -1/3$ suggesting more persistence in the former time series in a stronger sense than given. Figure 2 shows the measure of nonlinearity across the frequency domain. There is more nonlinearity at lower frequencies for $\theta = -1/3$ than $\theta = 1$ but this ranking is reversed at high frequencies.

Using this simple nonlinear moving average example we can consider the results of using other approaches on the GI. As suggested by a referee averaging the GI across histories gives:

$$E[GI_0(V_t, \mathcal{F}_{t-1})||V_t] = V_t$$

$$E[GI_1(V_t, \mathcal{F}_{t-1})||V_t] = \theta V_t$$

This produces results similar to the GI for a linear moving average of order 1 with parameters θ, σ_v^2 but with an autocovariance function different from the underlying time series. Alternatively we can use the collection of autocovariance functions derived above to find a Wold Representation (i.e., in this case a linear moving average of order 1) for each possible history:

$$\begin{aligned}
\phi_{v_{t-1}} &= \frac{1}{\theta V_{t-1}^2},\ \sigma_{v_{t-1}}^2 = \theta^2\ V_{t-1}^4 \quad \text{if } |\theta V_{t-1}^2| > 1, \\
\phi_{v_{t-1}} &= \theta V_{t-1}^2,\ \sigma_{v_{t-1}}^2 = 1 \quad \text{if } \ |\theta V_{t-1}^2| \leq 1.
\end{aligned}$$

These results can be compared to the Wold Representation of the time series itself which can be obtained from its autocovariance function. This gives:

$$\phi = \frac{(1+3\theta^2) \pm \sqrt[2]{(1+3\theta^2)^2 - 4\theta^2}}{2\theta}.$$

Consider the case $\theta = -1$. We know from the spectral density at zero of $X_t = X_{t-1} + Y_t$ that X_t is an integrated time series of order 1. However, using the definition of the GI that averages across history the long run response of X_t to shocks would be zero. The long response given by the Wold Representation of the time series is $1 + \phi = -1 + \sqrt{3} = 0.732$ and from the average of Wold Representation derived from the GI $1 + E[\phi_{v_{t-1}}] = 1 - 0.366 = 0.634$.

6 Computational Techniques

The example above extends to the whole class of nonlinear moving averages in cases where the moments of the innovation series are easy to find. However, it is not particularly useful in practice because of the difficulty of estimating such models. Nonlinear autoregressions are obviously easier to estimate but produce a large computational burden. As discussed in KPP it is possible to use Monte Carlo integration techniques to find a realization of GI_n.[¶] Essentially one just simulates a large number of realizations of the time series with two sets of initial conditions and average the difference. In order to estimate the measures provided here one needs to obtain many realizations of the GI using Monte Carlo integration for a range of histories. If we signify a collection of realizations of the GI for history $y^{t-1}(\omega)$ by $\{\widehat{gi}_n^i(y^{t-1}) : i = 1, \ldots, I\}$, then one can form

$$\frac{1}{I}\sum_{i=1}^{I}\widehat{gi}_{n+j}^i(y^{t-1})\widehat{gi}_n^i(y^{t-1})$$

to estimate $E_y^{t-1}(\omega)[GI_n + jGI_n]$. Note that since

$$\widehat{gi}_n^i(y^{t-1})\tilde{}N(GI(V_t,y^{t-1}(\omega)), \sigma_n^2(V_t, y^{t-1}(\omega))/M),$$

where M is the number of Monte Carlo replications used to produce $\widehat{gi}_n^i(y^{t-1})$ and $\sigma_n^2(V_t, y^{t-1}(\omega))$ is the variance across the realizations used to estimate the GI, the average of the crossproducts requires bias correction for $j = 0$. One can estimate the variance across the realizations during the calculation of $\widehat{gi}_n^i(y^{t-1})$ and estimate $E_y^{t-1}(\omega)[GI_n^2]$ by

$$\frac{1}{I}\sum_{i=1}^{I}\widehat{gi}_n^i(y^{t-1})\widehat{gi}_n^i(y^{t-1}) - \widehat{\sigma}_n^2(V_t, y^{t-1}(\omega))/M$$

For $j \neq 0$ the bias will go to zero as I increases and no bias correction is required.

Obviously it is not possible to compute this measure for each possible history if Y_t has a continuous distribution. Instead one might want to calculate for particular subsets of the histories. For example, one could define recession and expansion histories and compute the autocovariance functions for these two different histories.

The computational burden is substantially smaller for finite Markov chains where exact results are available. If $\mathbf{P}$ is the probability transition matrix of the Markov chain and $\mathbf{s}$ is the $k \times 1$ vector of values for the state then we have the following expression for the GI if at time $t-1$ the state is j and if at time t the state is i:

$$GI_n(i,j) = \mathbf{s}'\mathbf{P}^{n\prime}(e_i - \mathbf{P}'e_j),$$

where e_ℓ is a column vector of size k with all entries zero except for the ℓ-th row which contains a 1. If we define $v_{ij} = (e_i - \mathbf{P}'e_j)$ and use standard properties of Kronecker products we have

$$GI_n(i,j)GI_m(i,j) = v_{ij}'\left[\mathbf{s}'\mathbf{P}^{n\prime} \otimes \mathbf{P}^m\mathbf{s}\right]v_{ij.}$$

Thus, the main object that must be calculated is $[s'\mathbf{P}^{n\prime} \otimes \mathbf{P}^m s]$, if we denote it $\mathbf{B}(n,m)$ then we have:

$$E\left[GI_n(s_t, s_{t-1} = i)GI_m(s_t, s_{t-1} = i)\right] = \sum_{j=1}^{k} v_{ij}'\mathbf{B}(n,m)v_{ij}p_{ij.}$$

[¶] In this section I do not consider the issue of parameter uncertainty.

Since $\mathbf{B}(n,m)$ can be calculated in a simple recursive manner.

These results on Markov chains can be extended to the case of Markov switching models (see Hamilton 1989). Start by considering an observed time series Y_t that is a Markovian mixture of K independent unobserved time series $\{X_{kt}\}$ with $Y_t = X_{kt}$ if $S_t = k$.

Define $b_{kt} = P[S_t = k|\mathcal{F}_t]$ and $\widehat{b}_{kt} = P[S_t = k|\mathcal{F}_{t-1}]$, where $\mathcal{F}_t$ is generated only by the history of Y_t. In the expressions for directly observed Markov chains above we can replace v_{ij} with

$$\mathbf{b}_t - \widehat{\mathbf{b}}_t,$$

where $\mathbf{b}_t = [b_{1t}, \cdots, b_{Kt}]'$ and $\mathbf{b}_t = [b_{1t}, \cdots, b_{Kt}]'$. These vectors are produced as part of the estimation of the model for the observed sample. Finally the vector s is given by $[E[X_{1t}], \cdots, E[X_{Kt}]]'$. In order to simulate the GI and find average crossproducts of the GI one needs to obtain draws of the random vector $\mathbf{b}_t - \widehat{\mathbf{b}}_t$, which itself requires draws from the distributions of $\{X_{kt}\}$ and S_t.

The case of Markov switching models where the observed time series is the sum of two unobserved stochastic processes one of which is a Markov chain presents additional challenges. Suppose $Y_t = S_t + X_t$ where $S_t$is a Markov chain and X_t is a Gaussian autoregression with lag polynomial $\xi(L)$ and innovation V_t. Then one can redefine the states of the Markov chain such that $Y_t = \xi(L)Y_{t-1} + S_t^* + V_t$, with $S_t^* = S_t - \xi(L)S_{t-1}$. Then one can show that

$$GI_0 = Y_t - \mathbf{s}'\widehat{b}_t = (e_i - b_t) + \mathbf{s}'(\mathbf{b}_t - \widehat{\mathbf{b}}_t) + V_t,$$

and

$$GI_n = \psi_n GI_0 + \mathbf{s}'\mathbf{P}^{n\prime}(\mathbf{b}_t - \widehat{\mathbf{b}}_t),$$

where $\psi(L)(1 + \xi(L) = 1$.

The calculation of the measures of nonlinearity discussed in Section 4 is very difficult except in the finite Markov chain case. In that case we have:

$$NL(\lambda) = -\sum_{k=1}^{K} \ln(\frac{s(\lambda|y_{t-1} = k)}{\overline{s(\lambda)}})\ \pi(y_{t-1} = k),$$

where $\pi(y_{kt-1})$ is the unconditional probability of state k. Given $NL(\lambda)$ and $\overline{s(\lambda)}$ standard numerical techniques can be used to find the overall measure of nonlinearity. Once again if one is prepared to limit attention to subsets of history for continuously distributed random variables then a restricted version of the nonlinearity measure could be calculated. Of course, it will be smaller than one calculated for a finer information set. However, for comparison across models of the same time series one could choose the same subsets. One difficulty when simulation is used is that part of the measured nonlinearity will be just be simulation error. This effect can be reduced by simulating the underlying GI using the same future shocks for the pairs of realizations of the time series. Thus, if a time series is linear a GI constructed from just one realization will give the exact spectral density.

7 Empirical Examples

This Section is not finished

The construction of the GI and the properties of its autocovariance function are illustrated for two examples from the literature on nonlinearities in US GNP. Firstly for the Markov switching model of Hamilton (1989). In the notation from the previous section this model can be written:

$$Y_t = \xi_1 Y_{t-1} + \cdots + \xi_4 Y_{t-4} + S_t^* + V_t,$$

where $S_t^* = S_t - \xi(L)S_{t-1}$, and S_t is a two state Markov chain with $s_1 = \mu_1$ and $s_2 = \mu_2$ and $p = P[S_t = \mu_1 | S_{t-1} = \mu_1], q = P[S_t = \mu_2 | S_{t-1} = \mu_2]$.

If $\xi(L) = 0$ then the calculations simplify greatly. The hidden Markov chain now only has two states which without loss of generality we can consider to have values $1, 0$. Then for $n \geq 1$

$$E[Y_{t+n}|y^t(\omega)] - E[Y_{t+n}|y^{t-1}(\omega)] = (p+q-1)^n(b_{1t} - \widehat{b}_{1t}),$$

and the observed time series is linear heteroscedastic.

The second example is the two regime SETAR model of Potter (1995). In this case it is necessary to focus on subsets of the history of the time series. In the SETAR case it is natural to focus on the subsets of the history which define the regimes.

8 Conclusions

The characterization of the dynamics of nonlinear time series presented here is attractive because it allows the use of many linear techniques conditional on the history of the time series and integrating out the effect of the shock. Further, it shows that some definitions of the average dynamics of a nonlinear time series model can be obtained from its autocovariance function. However, just as in the case of linear time series model with non-Gaussian innovations, second order analysis does not provide a full characterization of the dynamics. For example, the methods of Potter (1995) allow one to assess the effect of negative and positive shocks for different histories that are hidden by the methods presented here.

Appendix

Proof of Theorem 1 *We can represent the time series in terms of a finite number of its GI and one conditional expectation by:*

$$Y_t = \sum_{n=0}^{N} GI_n(V_{t-n}, \mathcal{F}_{t-n-1}) + E[Y_t||\mathcal{F}_{t-n-1}],$$

since $Y_t = E[Y_t|\mathcal{F}_t]$. *Thus, we have:*

$$\{\sum_{n=0}^{N} GI_n(V_{t-n}, \mathcal{F}_{t-n-1})\}^2 \leq Y_t^2 + 2|Y_t E[Y_t||\mathcal{F}_{t-n-1}]| + (E[Y_t||\mathcal{F}_{t-n-1}])^2.$$

Hence, if the time series is square integrable the rhs of this expression can be bounded by an integrable function and by the dominated convergence theorem the expectation of $\{\sum_{n=0}^{\infty} GI_n(V_{t-n}, F_{t-n})\}^2$ *is finite. Applying the expectation operator to the square of the finite sum and making the usual expansion we have*

$$\sum_{n=0}^{N} E[GI_n(V_{t-n}, \mathcal{F}_{t-n})^2] + 2\frac{N}{N+1}\sum_{n=1}^{N} E[GI_n(V_{t-n}, \mathcal{F}_{t-n-1})GI_{n-1}(V_{t-n-1}, \mathcal{F}_{t-n-2})] +$$
$$\cdots + 2E[GI_0(V_t, \mathcal{F}_{t-1})GI_N(V_{t-N}, \mathcal{F}_{t-N-1})].$$

By the properties of the GI only the expectation of terms of the first sum have a none zero value and we can conclude that

$$E[Y_t^2] = \sum_{n=0}^{\infty} E[GI_n(V_{t-n}, \mathcal{F}_{t-n})^2].$$

Using the property that for $n < 0$, $GI_n(V_t, F_{t-1}) = 0$ *completes the proof.*

Proof of Theorem 2 *Signify the elements of* $\mathcal{A}_H$ *by* $c_{y^{t-1}(\omega)}$. *In order to establish that a function defined on the integers is autocovariance function we need to show the following:*

1. $c_{y^{t-1}(\omega)}(0) \geq 0$.
2. $c_{y^{t-1}(\omega)}(0) \geq |c_{y^{t-1}(\omega)}(j)|$ *for* $j = \pm 1, \pm 2, \ldots$.
3. $c_{y^{t-1}(\omega)}(j) = c_{y^{t-1}(\omega)}(-j)$ *for* $j = \pm 1, \pm 2, \ldots$
4. $c_{y^{t-1}(\omega)}(j)$ *is non-negative definite function on the integers. A function* κ *on the integers is non-negative definite if and only if:*

$$\sum_{i,\ell=1}^{m} a_i \kappa(j_i, j_\ell) a_\ell \geq 0.$$

The only condition of 1. to 4. that is not a direct property of the inner product $\langle\cdot,\cdot\rangle_{y^{t-1}(\omega)} = \sum_{n=-\infty}^{\infty} E_{y^{t-1}(\omega)}[Z_n X_n$ *on* $\mathcal{A}$ *is 4 (see Dunford and Schwartz (1957 IV.4.19). In order to show this condition. Define:*

$$\mathbf{a} = (a_1, a_2, \ldots, a_m)',\ \mathbf{j} = (j_1, j_2, \ldots, j_m) \in \textit{integers},\ \mathbf{X^J}_n = (Z_{n+j_1}, Z_{n+j_2}, \ldots, Z_{n+j_m}).$$

Consider:

$$\begin{aligned}
&\sum_{i,\ell=1}^{m} a_i c_{y^{t-1}(\omega)}(j_i - j_\ell) a_\ell = \\
&\quad \sum_{i,\ell=1}^{m} a_i \langle \mathbf{Z}_{j_i}, \mathbf{Z}_{j_\ell} \rangle_H a_\ell = \\
&\quad \sum_{i,\ell=1}^{m} \langle a_i \mathbf{Z}_{j_i}, a_\ell \mathbf{Z}_{j_\ell} \rangle_H = \\
&\qquad\qquad \langle \sum_i a_i \mathbf{Z}_{j_i}, \sum_\ell a_\ell \mathbf{Z}_{j_\ell} \rangle_H.
\end{aligned}$$

Using the definition of the inner product consider one term in the infinite sum over n:

$$E_{y^{t-1}(\omega)}[\sum_i a_i Z_{n+j_i}, \sum_\ell a_\ell Z_{n+j_\ell}] = E_{y^{t-1}(\omega)}[\mathbf{a}' \mathbf{X^j}_n \mathbf{X^j}'_n \mathbf{a}] \geq 0.$$

Thus, $c_{y^{t-1}(\omega)}(j)$ *is a non-negative definite function.*

Proof of Corollary to Theorem 2 *Note that a time series* Y_t *can be expressed as:*

$$Y_t = \sum_{n \geq 0} GI_n(V_{t-n}, \mathcal{F}_{t-n-1}),$$

since the lead term in the sum is $E[Y_t || V_t, \mathcal{F}_{t-1}] = Y_t$ *and all the other terms cancel.*

Without loss of generality assume that Y_t *has mean zero, then*

$$\gamma(k) = E[Y_{t+k} Y_t] = E[\sum_{n \geq 0} GI_n(V_{t+k-n}, \mathcal{F}_{t+k-n-1}) \sum_{n \geq 0} GI_n(V_{t-n}, \mathcal{F}_{t-n-1})].$$

This expression can be split into separate terms of form:

$$E[GI_\ell(V_{t+k-\ell}, \mathcal{F}_{t+k-\ell-1}) GI_m(V_{t-m}, \mathcal{F}_{t-m-1})].$$

Note that if $k - \ell \neq m$ *then by the law of iterated expectations the term has value zero. Since if* $k - \ell < m$ *one can condition on information up to time* $t - m$ *and if* $k - \ell > m$ *one can condition on information up to* $t - k - \ell$. *In the cases where* $k - \ell = m$ *we have:*

$$E\left[E\left[GI_\ell(V_{t-m}, \mathcal{F}_{t-m-1}) GI_m(V_{t-m}, \mathcal{F}_{t-m-1}) | y^{t-m-1}(\omega)\right]\right].$$

Collecting all similar terms proves the result.

References

[1] Billingsley, P., (1986), *Probability and Measure*, New York: John Wiley.

[2] Brillinger, D. (1981),

[3] Brockwell and Davis, (1987) *Times Series: Theory and Methods*, New York:

[4] Cochrane, J., (1988) How big is the random walk in US GNP? *Journal of Political Economy,96, 893-920.*

[5] Dunford and Schwartz (1957), *Linear Operators Part 1*, New York: John Wiley.

[6] Gallant, A.R., P.E. Rossi and G. Tauchen (1993). "Nonlinear Dynamic Structures," *Econometrica*, 61, 871-908.

[7] Koop, G., Pesaran, H., and Potter, S., (1995): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics,* 74 119-147.

[8] Potter, S., (1995), "Nonlinear Impulse Response Functions," *mimeo,* Department of Economics, University of California, Los Angeles.

[9] Priestley, M.B., (1988), *Non-Linear and Non-Stationary Time Series,* New York: Academic Press.

[10] Rosenblatt, M., (1971), *Markov Processes. Structure and Asymptotic Behavior,* New York: Springer.

[11] Tong, H., (1990), *Non-linear Time Series: A Dynamical System Approach,* Oxford: Clarendon Press.

[12] Tong, H., (1995), "A personal overview of nonlinear time series analysis from a chaos perspective," *Scandavian Journal of Statistics,* forthcoming.

[13] Tweedie, R., (1983), "The Existence of Moments for Stationary Markov Chains," *Journal Of Applied Probability*, 20, 191-6.